



Kanton Zürich
Direktion der Justiz und des Innern
Statistisches Amt

R Austausch

25.07.2024

Miriam Hofstetter & Oliver Roos - AMOSA

Thomas Knecht & Philipp Bosch - Statistisches Amt



Agenda

- R @ AMOSA
 - Sequenzanalyse von Arbeitsverläufen mit R.
- useR!-Conference Highlights
 - Parallelisierung (futureverse)
 - I/O mit Rio
 - Httr2
 - R-Universe
- ChatGPT & R



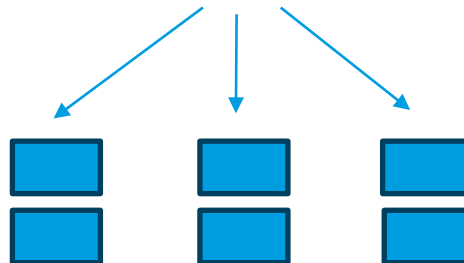
Parallelisierung 1

- Was löst die Parallelisierung
 - Zeitgewinn!

Seriell



Parallel





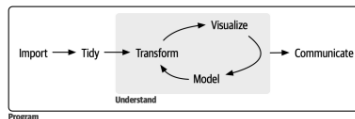
Parallelisierung 2

- Wann lohnt sich eine Parallelisierung:
 - Sobald ein rechenintensiver loop X-Fach ausgeführt wird
 - Prozess muss in sich geschlossen sein
 - Bei sehr kleinen und nicht rechenintensiven Prozessen ist der Zeitgewinn nicht sehr gross, da der Overhead für die Verteilung der Prozesse stärker ins Gewicht fällt



I/O Rio – Was löst das package?

Why you should care about I/O



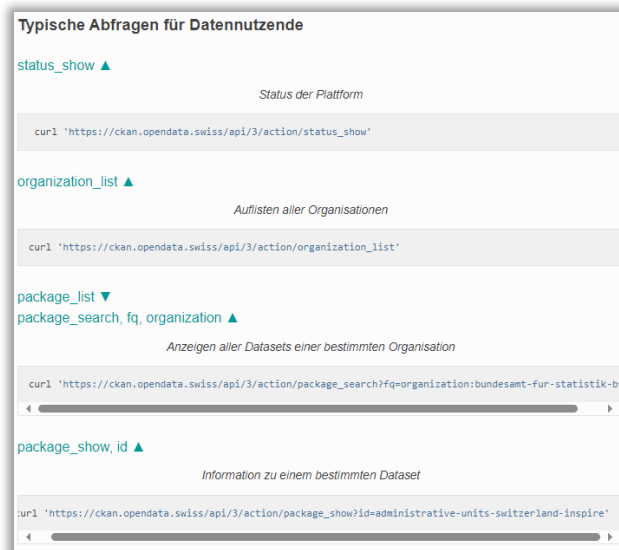
Source: R for Data Science

“First, you must import your data into R. This typically means that you take data stored in a file, database, or web application programming interface (API) and load it into a data frame in R. **If you can’t get your data into R, you can’t do data science on it!**”



HttR2 – Was löst das package?

- Manche Daten kann nicht mal Rio einlesen
 - Zum Beispiel hinter einer API






R-Univers – Wie hilft's mir?

What is R-universe?






- **Discover:** a *searchable catalogue* for R software, articles, datasets, anywhere in the ecosystem. Indexed and ranked using R specific features.
- **Learn:** provides rendered material and extensive background information to learn about a package and get started using it.
- **Publish:** an open platform for running personal cran-like package repositories (cf. KH keynote) to make your work visible and accessible to a big audience, with or without CRAN.
- **Develop:** a zero config development environment for actual continuous *integration*, experimental R features, new platforms, etc.

ChatGPT

- Produkt der US Firma OpenAI
- Large Language Model (LLM) = neuronales Netz, trainiert mit riesigen Mengen an Textdaten
- Im Prinzip «Satzvervollständigung» → Was ist das wahrscheinlichste nächste Wort?
- Wichtiges Vokabular: «Prompt» & «prompting»
 - Die Eingabeaufforderung also den Text oder die Frage welche man ChatGPT stellt 



Was gilt für uns im Kanton?

- [Einsatz von ChatGPT & Co. in der Verwaltung \(ktzh.ch\)](https://ktzh.ch)
 -  KEINE Personendaten
 -  KEINE Daten welche Rückschluss auf Amt/Direktion erlauben
 -  KEINE internen, vertraulichen oder geheimen Daten
 -  KEINE (Ausschnitte) von Anfragen aus der Bevölkerung
 -  Öffentlich Verfügbare Informationen & **Daten (=OGD)**

ChatGPT

Takeaways

- Extrem hilfreich, wenn man das Endprodukt und die prinzipiellen Schritte kennt
- So spezifisch wie möglich im Prompt sein
 - Nennung von packages & gewünschter Syntax
 - Herunterbrechen in Schritte
 - Fehlermeldung mit Hinweis in Prompt
- Better save than sorry was Datenschutz & Compliance angeht!

Neues zur Standardapplikation

- Fachgruppe gebildet und am 19.09 Kick-Off Meeting
- Es werden mir nur noch wenige Tickets vom AFI weitergeleitet :)
- Wichtig:
 - Neues Update in die Wege leiten!
 - Backlog erstellen

Wenn ihr also Rückmeldungen habt
gerne bei uns melden!

